Available online at www.sciencedirect.com

SCIENCE DIRECT®

Remote Sensing of Environment 91 (2004) 452–468

Remote Sensing of Environment

www.elsevier.com/locate/rse

Thematic accuracy of the 1992 National Land-Cover Data for the western United States

J.D. Wickham[a,*], S.V. Stehman[b], J.H. Smith[c], L. Yang[d]

[a] *National Exposure Research Laboratory, US Environmental Protection Agency (E243-05), Research Triangle Park, NC 27711, USA*
[b] *SUNY-ESF, 320 Bray Hall, Syracuse, NY 13210, USA*
[c] *Geographic Analysis and Monitoring Program, US Geological Survey, National Center, Mail Stop 519, Reston, VA 20192, USA*
[d] *Science Application International Corporation (SAIC), EROS Data Center, Sioux Falls, SD 57198, USA*

Received 3 November 2003; received in revised form 16 April 2004; accepted 18 April 2004

Abstract

The MultiResolution Land Characteristics (MRLC) consortium sponsored production of the National Land Cover Data (NLCD) for the conterminous United States, using Landsat imagery collected on a target year of 1992 (1992 NLCD). Here we report the thematic accuracy of the 1992 NLCD for the six western mapping regions. Reference data were collected in each region for a probability sample of pixels stratified by map land-cover class. Results are reported for each of the six mapping regions with agreement defined as a match between the primary or alternate reference land-cover label and a mode class of the mapped 3×3 block of pixels centered on the sample pixel. Overall accuracy at Anderson Level II was low and variable across the regions, ranging from 38% for the Midwest to 70% for the Southwest. Overall accuracy at Anderson Level I was higher and more consistent across the regions, ranging from 82% to 85% for five of the six regions, but only 74% for the South-central region.

Keywords: Land-cover mapping; MRLC; NLCD; Thematic accuracy assessment

1. Introduction

In 1992, a group of federal agencies formed the Multi-Resolution Land Characteristics (MRLC) consortium (Loveland & Shaw, 1996) to meet growing needs for satellite-based land-cover information amid the rising costs of Landsat data (Goward, 1989). Federal agencies were able to offset rising costs by pooling resources and operating as a consortium.

Several mapping projects utilized the satellite data acquired through the MRLC consortium, including the USGS GAP habitat mapping project (Scott et al., 1993, Scott & Jennings, 1998), the NOAA C-CAP coastal change program (Dobson et al., 1995), and the National Land Cover Data (NLCD) mapping project (Vogelmann et al., 2001). The primary objectives of the NLCD project were to map general land-cover categories for the conterminous United States and document the map's thematic accuracy. A 21-class legend was adopted that followed the Anderson et al. (1976) classification system. Descriptions of the mapping procedures and protocols can be found in Vogelmann et al. (1997, 1998a, 1998b) and internet documentation maintained by the USGS EROS Data Center (http://landcover.usgs.gov/natllandcover.html). The target year for satellite acquisition was 1992, although dates ranged from 1986 to 1996 due to lack of suitable imagery.

NLCD mapping of the conterminous United States was completed late in the 2000 calendar year. (Vogelmann et al., 2001). Thematic accuracy assessments have followed mapping. Thematic accuracy results for the four mapping regions of the eastern U.S. were reported in Stehman et al. (2003). Here, we report thematic accuracy results for the six western mapping regions of the conterminous United States. These results combined with those reported in Stehman et al. (2003) complete a national accuracy assessment of the 1992 NLCD.

2. Methods

2.1. Sampling design

Complete details of the sampling design, response design, and analysis components of the NLCD accuracy

* Corresponding author. Tel.: +1-919-541-3077; fax: +1-919-541-1138.
E-mail address: wickham.james@epamail.epa.gov (J.D. Wickham).

0034-4257/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.rse.2004.04.002

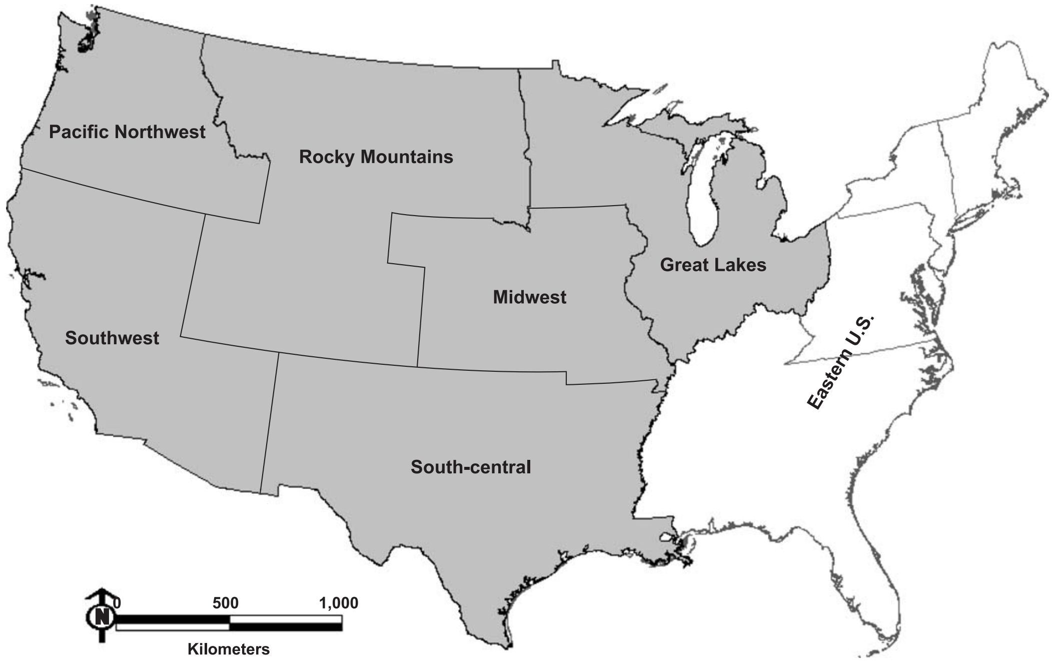

Fig. 1. Location map. Accuracy estimates are reported for the mapping regions shaded in gray. The eastern U.S. includes the Southeast, mid-Atlantic, New York and New Jersey, and New England mapping regions.

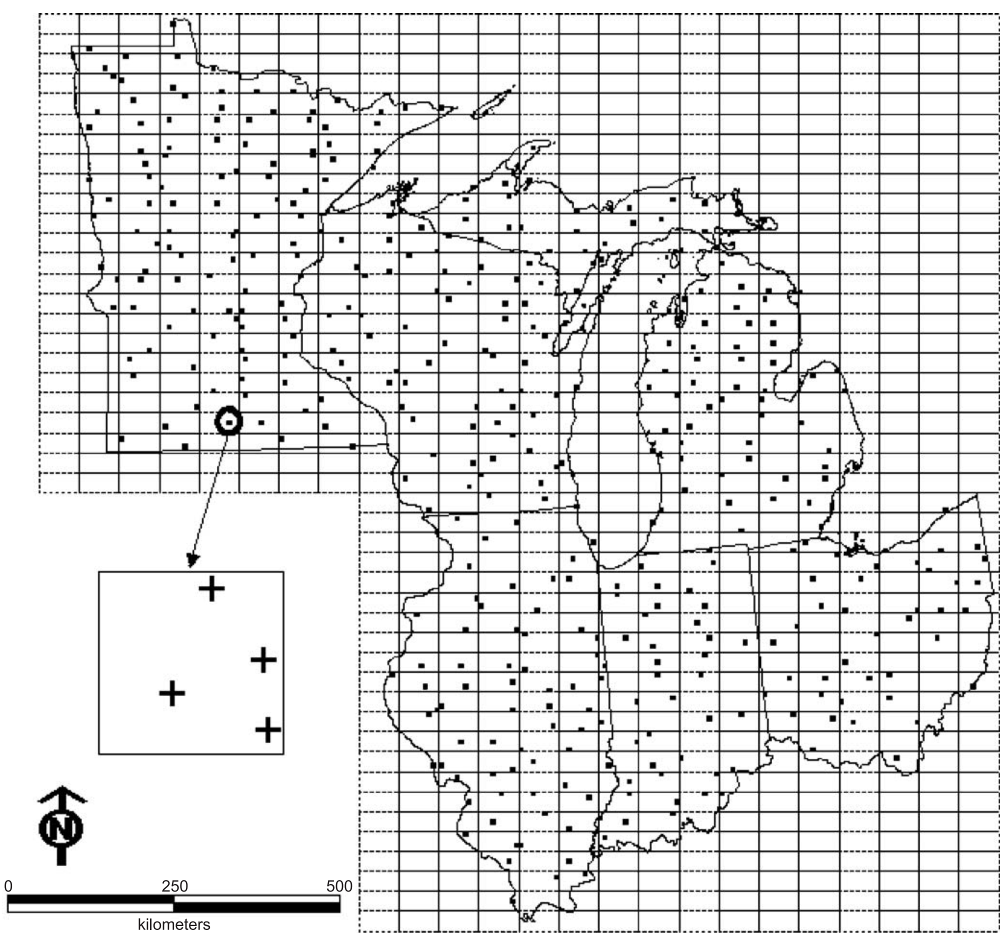

Fig. 2. Two-stage cluster design. Dashed lines are the tessellation cells, and solid-filled squares are the selected PSUs. Sample pixels are shown as crosshairs in "exploded" PSU. Only PSUs that included sampled pixels are shown (some tessellation cells do not include a solid-filled square).

assessments are reported in Stehman et al. (2003). We provide a synopsis. The sampling design incorporated a two-stage cluster sampling protocol and three levels of stratification. The six separate mapping regions of the western U.S. formed the first level of stratification (Fig. 1). The second level of stratification was a tessellation of each mapping region into cells of 60×30 km. Each tessellation cell was further subdivided into 6×6 km primary sampling units (PSU), with each PSU occupying 2% of the tessellation cell. These definitions of tessellation cells and PSUs differ somewhat from the design used for the four mapping regions in the eastern U.S. (Stehman et al., 2003), where National Aerial Photography Program (NAPP) photography frames formed the PSUs and either 15×15 or 30×30 min cells formed the tessellation cells.

One PSU from each tessellation cell was selected at random, representing the first stage of the cluster sampling design. Once sample PSUs were selected from the tessellation cells, all pixels within these sampled PSUs were assigned to strata based on the map land-cover class (the third level of stratification). The full sample consisting of 100 pixels from each class mapped in a region was then obtained via stratified random sampling from the pixels available in the first-stage sample PSUs (Fig. 2). The selected pixels represent the second stage of the cluster sampling design, and are referred to as secondary sampling units (SSU). The number of classes mapped in the six western regions ranged from 18 to 21 (Table 1), resulting in a total of 11,900 sample pixels for all six mapping regions.

We chose a per-class sample size of 100 (per region) because it results in an expected standard error for user's accuracy of 0.05 for simple random sampling within a stratum, if the true user's accuracy is 0.50. The realized standard errors will differ from 0.05 for a variety of reasons. For example, the clustering in the NLCD design will tend to increase the standard error relative to simple random sampling. Conversely, user's accuracy will rarely be 0.50, and the standard error will decrease as the true user's accuracy deviates from 0.50. Selection of a sample size of 100 reflected consideration of both target precision of the user's accuracy estimates and overall data collection costs.

For the Pacific Northwest region, a supplemental sample was selected to increase the sample size for three rare classes: perennial snow or ice (12), high-density residential (22), and mining (32). Each of these classes possessed a high degree of clustering of pixels into a few first-stage PSUs, so the supplemental sample was included to improve the spatial distribution of the sample for these classes. Twenty-five additional pixels were selected by simple random sampling from all pixels of that class, not just those in the first-stage PSUs. The supplemental sample increased the total number of sampled pixels to 11,975 in the six western regions.

2.2. Response design

NAPP photography, Digital Orthophoto Quarter Quadrangles (DOQQ), and other high-resolution aerial photographic sources were interpreted to obtain reference land-cover labels. DOQQs are scanned aerial photographs

Table 1
Regional distribution of mapped land cover (% of area) for the 1992 NLCD Level II and Level I classifications

Class names	Class codes		Class areal percentages by mapping region					
	Level II	Level I	GL	SC	MW	RM	SW	PNW
Water	11	10	5.209	3.507	1.127	1.377	0.768	2.183
Perennial snow/ice	12	10	0.000	0.000	0.000	0.057	0.005	0.096
Low-density residential	21	20	1.397	0.675	0.520	0.203	0.936	0.595
High-density residential	22	20	0.592	0.250	0.226	0.027	0.118	0.004
Commercial/transportation	23	20	0.839	0.502	0.660	0.197	0.443	0.416
Bare rock, sand, clay	31	30	0.022	0.812	0.096	2.331	5.002	1.775
Mining	32	30	0.124	0.057	0.034	0.042	0.068	0.011
Transitional	33	30	0.214	0.296	0.022	0.347	0.035	1.521
Upland deciduous forest	41	40	21.390	8.337	10.435	2.300	.841	2.452
Upland evergreen forest	42	40	2.967	8.978	0.637	13.798	16.620	36.384
Upland mixed forest	43	40	2.829	3.577	0.870	0.307	1.466	2.732
Shrubland	51	50	0.133	22.718	0.388	20.842	57.452	29.868
Orchards and vineyards	61	80	0.000	0.008	0.000	0.000	1.100	0.250
"Semi-natural" Grassland	71	50	0.627	22.694	27.691	35.019	10.074	8.627
Pasture/hay	81	80	15.809	9.838	16.736	4.403	2.151	4.552
Row crops	82	80	36.781	9.352	30.264	7.378	1.635	1.480
Small Grains	83	80	0.665	4.410	7.532	6.818	0.945	4.117
Fallow	84	80	0.000	0.046	0.553	3.468	0.011	2.233
Urban/recreational Grasses	85	80	0.517	0.120	0.272	0.049	0.093	0.046
Woody wetlands	91	90	7.623	2.024	0.892	0.196	0.050	0.220
Emergent wetland	92	90	2.262	1.799	1.046	0.839	0.190	0.440

All pixels within each mapping region were classified. Mapping regions are abbreviated GL (Great Lakes), SC (South-central), MW (Midwest), RM (Rocky Mountain), SW (Southwest), and PNW (Pacific Northwest).

corrected for image displacement and camera tilt. DOQQ resolution is 1 m covering an area of 3.75 × 3.75 min, or one-quarter of the area of a 1:24,000 scale USGS quadrangle. The DOQQs were primarily used in the South-central and Midwest regions, and were almost exclusively black-and-white. High-resolution aerial photographic sources were used primarily in the Rocky Mountain and Southwest regions. NAPP photography was used in the Great Lakes and Pacific Northwest, as well as the Rocky Mountain and Southwest when high-resolution photographic sources were not available. The reference sample was selected independently of the locations used to develop the NLCD classification.

Coordinates for each sample pixel were overlaid digitally on a color composite of Landsat imagery and then located by context on the aerial photography. When DOQQs were used, sample pixels were overlaid digitally on both the Landsat and DOQQs. Reference land-cover labels were identified by visual interpretation of the aerial photography. Photointerpretation was done using a blind format in that the photointerpreters did not have knowledge of the map classification. Data collected through photointerpretation included primary and alternate land-cover labels, nominally scored interpretation and location confidence ratings (ICR and LCR), a relative location score and other information (Table 2). Use of alternate reference labels was intended for those situations where a single label could not be assigned with confidence. Confidence and location ratings were collected to provide additional information on reference data quality and to provide options for alternative analyses of the data (e.g., analyses by subsets defined by confidence ratings).

Table 2
Attribute information collected for interpreting agreement between map and reference data

Information from Reference Source
(1) Primary reference label: land-cover label thought to be most correct by photointerpreter.
(2) Alternate reference label: land-cover label that might also be considered correct given information in the photograph. An alternate reference label was not provided if the photointerpreter's primary reference label was believed unambiguously evident.
(3) Photointerpreter confidence (ICR) : nominal ranking of photointerpreter's certainty in identifying reference land-cover label. Rank values range from 1 (not confident) to 4 (very confident). ICR values for Great Lakes were assigned values ranging from 1 (doubtful) to 3 (absolutely correct).
(4) Location confidence (LCR): nominal ranking of photointerpreter's certainty in identifying the correct location on the reference medium. Rank values range from 1 (not confident) to 4 (very confident). LCR values for Great Lakes were assigned values ranging from 1 (doubtful) to 3 (absolutely correct).
(5) Relative location: nominal score for location of sample pixel: 1 = on the edge between two land-cover classes; 2 = homogeneous area of land cover, and; 3 = heterogeneous area of land cover. Relative location scores were not collected in the Great Lakes.
(6) Date: day, month, and year of photo acquisition.
(7) Notes: entries for any other factors that may affect photointerpretation (e.g., temporal change).
Information from Map Source
(1) Center: land-cover label of pixel selected as a sample.
(2) Modal: most frequent land-cover label in a 3 × 3 pixel neighborhood surrounding selected sample pixel.
(3) Any: list of all land-cover labels in a 3 × 3 pixel neighborhood surrounding selected sample pixel.

Three different teams were involved in reference data collection for the six western mapping regions, and each individual mapping region was assigned to only one team. Each team conducted photointerpretation training prior to collection of reference information for each mapping region. The purpose of the training was to reduce individual subjectivity in reference label assignment. The photointerpretation teams also assigned a project manager, who was typically the most experienced photointerpreter. The project manager randomly checked a subsample of reference labels assigned by each photointerpreter, and precedence was given to the project manager's assignment when they disagreed.

Map label information included the land-cover label at the sample (center) pixel, a list of all land-cover labels in a 3 × 3 pixel neighborhood centered on the selected sample pixel, and the modal values for the 3 × 3 neighborhood (Table 2). The variety of reference and map label data supported several different agreement definitions (Stehman & Czaplewski, 1998) that can be used to produce accuracy estimates that differ in their sensitivity to errors in the reference data (Hammond & Verbyla, 1996; Verbyla & Hammond, 1995). For example, restricting agreement to be a match between the center (map) pixel label and only the primary reference label regardless of interpretation confidence rating may overestimate true classification error by incorporating reference label measurement error related to difficulty in identifying the correct class or the correct location on the photo. Conversely, defining agreement as a match between any map label in the 3 × 3 window and either the primary or alternate label only for sites whose interpretation confidence is high may underestimate true classification error. For consistency with previous NLCD accuracy assessment reports (Yang et al., 2001; Stehman et al., 2003), we defined agreement as a match between the primary or alternate reference label and the modal land-cover label of the 3 × 3 pixel neighborhood from the map, regardless of ICR or LCR scores. If there was more than one modal value (e.g., two land-cover labels each occupied 4 of the 9 pixels in the 3 × 3 neighborhood), a match between the primary or alternate reference label and either modal value was considered agreement.

To illustrate the effect of agreement definition on accuracy, we also report estimated accuracy for three other definitions of agreement for the three regions having the highest number of sample pixels for which an alternate reference land-cover label was assigned. The first definition ("CenterP") defines agreement as a match between the map class of the sample pixel and the primary reference label only. This definition makes no allowance for the co-location error of the reference and map pixel or for ambiguity in the reference land-cover label. The second definition ("Center") defines agreement as a match between the map class of the sample pixel and either the primary or alternate reference label. This definition also does not accommodate colocation error, but including the alternate reference label makes some allowance for pixels that cannot be unambiguously assigned to a single class. The third definition ("Homogeneous") includes only the subset of the sample in which all pixels within the 3×3 window centered on the sample pixel have the same map land-cover class. Agreement is defined as a match between the map label and either the primary or alternate reference label. The "Homogeneous" definition largely removes the spatial registration problem because it considers only pixels centered within homogeneous areas defined by the mapped land cover, and it accommodates ambiguity in the reference land-cover label.

2.3. Analysis

The cell proportions of the error matrix and accompanying accuracy measures were estimated following standard probability-sampling protocols to insure statistical consistency of the estimators. Estimates were derived by weighting each sample pixel by the inverse of its inclusion probability derived from the sampling design (Stehman, 2000; Stehman et al., 2003). For the three rare land-cover classes targeted by the supplemental sample for the Pacific Northwest region, the inclusion probabilities were calculated using the probability rule for the union of two independent events. Because the two samples were selected independently, the probability that a pixel is included in the combined sample is the sum of the inclusion probabilities from the two individual samples minus the product of the individual sample inclusion probabilities. The inclusion probabilities for all other land-cover classes were determined from the general design implemented in the region.

Reference labels were collected for 11,628 of the 11,900 (98%) sample pixels from the original design, and 60 of the 75 (80%) supplemental pixels added in the Pacific Northwest. Reference information was not collected for 287 sample pixels due to unavailable photos or other factors. No sample pixels were missing reference labels in the Midwest or South-central regions. Sample pixels for which a reference label was not obtained were treated as missing at random.

Alternate reference labels were recorded for 4,654 of the 11,688 interpreted sample pixels (40%). The proportion of alternate reference labels assigned was noticeably different among the three groups that collected the reference data. Interpretation confidence rating (ICR) scores of "confident" or "absolutely confident" were recorded for 10,536 of the 11,688 pixels with reference data (90%). About 64% (2974/4654) of the sample pixels that had an alternate reference label also had an ICR score of "absolutely confident". An ICR score of "absolutely confident" and identification of an alternate reference label are somewhat inconsistent since the option of assigning an alternate reference label was incorporated to address those situations where ambiguity prevented confident identification of a reference label. We used all reference data to construct the regional error matrices regardless of ICR score.

Accuracy estimates and their accompanying standard errors were computed using the SURVEYMEANS procedure in SAS (SAS, Version 8.2, 2001). A standard error may be used to form an approximate 95% confidence interval by adding and subtracting twice its value to the estimate of interest (user's, producer's, and overall accuracy). The standard errors were computed for the six western regions using the PSU as the cluster. In contrast, map polygons were used as the clusters for the standard error estimates in the four eastern U.S. mapping regions (Stehman et al., 2003). Using the PSUs as the clusters is more consistent with the actual design implemented. A more detailed discussion of the variance estimation methodology is provided in Stehman et al. (2003).

3. Results

3.1. Error matrices

Overall agreement scores at Anderson Level II ranged from 38% (Midwest) to 70% (Southwest) (Tables 3–8). The very low overall accuracy for the Midwest (Table 4) was due to considerable confusion between row crop (82), small grains (83), and fallow (84) classes. When these classes were combined into a single class (i.e., cropland), overall accuracy improved to 56%. Different types of cropland would not be distinguished until the fourth level of the Anderson et al. (1976, p. 10) hierarchy. Treating row crops, small grains, and fallow as a single class would improve Anderson Level II overall accuracy in the other five mapping regions by only about 5% or less.

All Level II error matrices revealed a consistent pattern in that the dominant sources of error were associated with the areally dominant land-cover classes. The frequency of non-zero cell entries was highest for some combination of shrubland (51), "semi-natural" grassland (71), and hay or pasture (81) in five the six mapping regions. Cropland (82) also had a high frequency of non-zero cell entries in the Great Lakes mapping region. For example, 25% of the map error in the Midwest was confusion of cropland with hay/pasture and small grains (Table 4), and over 10% of the map error in the Pacific Northwest was confusion between shrubland and "semi-natural" grassland (Table 8). Commission errors for "semi-natural" grassland were greater than 12% in four of the six mapping regions.

Table 3
Great Lakes, Anderson Level II

	11	21	22	23	31	32	33	41	42	43	51	71	81	82	83	85	91	92	Total	Users	S.E.	*n*
11	**5.741**	0.018								0.035	0.179								5.973	0.96	0.03	42
21		**1.418**	0.007	0.089				0.018		0.036	0.027	0.006		0.009		0.036	0.018		1.664	0.85	0.04	111
22		0.408	**0.161**	0.036						0.018				0.014		0.027			0.664	0.24	0.07	89
23	0.009	0.388	0.009	**0.502**	0.009						0.009			0.009	0.027	0.046	0.009		1.017	0.49	0.17	73
31	0.001				**0.005**									0.001					0.007	0.59	0.21	61
32	0.033		0.003	0.019		**0.083**	0.011	0.006		0.006	0.025	0.008	0.003	0.008	0.002	0.003	0.030	0.010	0.250	0.33	0.11	92
33	0.002			0.004	0.002	0.009	**0.083**	0.010	0.002	0.007	0.004	0.002	0.002	0.004	0.004		0.002	0.008	0.145	0.56	0.08	68
41	0.299	0.066		0.232	0.001		0.830	**13.789**	0.260	2.160	1.081	0.267	0.461	0.309			2.120	0.697	22.572	0.61	0.05	200
42				0.035		0.030		0.091	**1.606**	0.521	0.061			0.030			0.182		2.556	0.63	0.06	82
43							0.204	0.311	0.070	**1.995**							0.065	0.035	2.680	0.74	0.07	69
51		0.001		0.002		0.001	0.015	0.005		0.002	**0.043**	0.002		0.001	0.001		0.002	0.006	0.081	0.52	0.08	65
71		0.243			0.007	0.014	0.014	0.035	0.014	0.037	0.021	**0.267**	0.007	0.057		0.007	0.028	0.049	0.800	0.33	0.22	51
81		0.357		0.509	0.002	0.006	0.169	0.481		0.481	0.361	0.981	2.422	7.310	0.651	0.804	0.516	0.001	15.051	0.16	0.04	114
82		0.572		0.415	0.009			0.342		0.565	0.229	0.352	0.035	**29.990**		0.507	0.349		33.365	0.90	0.03	167
83		0.008		0.176				0.008				0.048	0.063	0.104	**0.763**	0.008	0.008	0.016	1.202	0.63	0.14	91
85	0.006	0.063		0.019	0.006	0.006	0.006	0.006	0.006	0.006		0.018		0.024		**0.246**		0.018	0.430	0.57	0.07	66
91	0.103						0.207	0.886	1.092	1.088	1.046						**3.754**	0.660	8.836	0.42	0.07	108
92	0.124			0.060	0.030		0.060	0.123	0.060	0.156	0.224	0.038		0.393	0.091		0.402	**0.937**	2.698	0.35	0.08	85
Total	6.138	3.542	0.180	2.098	0.071	0.149	1.599	16.111	3.110	7.113	3.310	1.989	2.993	38.263	1.539	1.684	7.485	2.437	99.991			
Prod.	0.91	0.40	0.89	0.24	0.06	0.55	0.05	0.86	0.52	0.28	0.01	0.13	0.81	0.78	0.50	0.15	0.50	0.38				
S.E.	0.04	0.08	0.07	0.07	0.05	0.22	0.02	0.03	0.08	0.05	0.00	0.11	0.10	0.03	0.15	0.05	0.07	0.10			0.02	
n	71	180	24	105	49	44	81	166	75	130	85	38	32	244	69	65	103	73				1634

Great Lakes, Anderson Level I[a]

	10	20	30	40	50	80	90	Total	Users	S.E.	n
10	**5.428**	0.018		0.035	0.179			5.660	0.96	0.03	37
20	0.009	**3.069**	0.009	0.036	0.048	0.137	0.018	3.327	0.92	0.02	273
30	0.034	0.027	**0.192**	0.030	0.034	0.025	0.051	0.393	0.49	0.07	217
40	0.297	0.336	1.003	**22.737**	0.845	0.800	1.956	27.974	0.81	0.03	363
50	0.001	0.018	0.051	0.086	**0.349**	0.066	0.070	0.641	0.54	0.17	112
80		1.450	0.195	1.414	1.452	**44.904**	0.974	50.389	0.89	0.02	441
90	0.196	0.060	0.298	3.328	1.102	0.484	**6.149**	11.617	0.53	0.05	191
Total	5.965	4.978	1.748	27.666	4.009	46.416	9.218	**100.000**			
Prod.	0.91	0.62	0.11	0.82	0.09	0.97	0.67				
S.E.	0.04	0.09	0.04	0.03	0.06	0.01	0.06			0.02	
n	66	307	172	381	112	429	167				1634

Error matrix for Level I and Level II classification: Great Lakes region. Level II overall accuracy was 64% with a SE of 2%, and Level I overall accuracy was 83% with a S.E. of 2%.

Reference land-cover labels form the columns and map land-cover labels form the rows. The value for each cell of the error matrix represents an estimated percentage area of the region. For example, a cell entry of 8.523 would translate to 8.523% of the region's area.

The column labeled "Users" is the estimated user's accuracy, and the row labeled "Prod." is the estimated producer's accuracy (Story & Congalton, 1986).

S.E. is the standard error of the user's (row) or producer's (column) accuracy estimate. The total number of samples for the row or column is labeled as "*n*."

Small discrepancies between actual and reported row or column totals are due to rounding errors. Cell proportions with the value zero (0) are left blank to assist visualization of the error patterns.

[a] Shrubland (51) and grassland (71) combined into single Anderson Level I class (50).

Table 4

Midwest, Anderson Level II

	11	21	22	23	31	32	33	41	42	43	51	71	81	82	83	84	85	91	92	Total	Users	S.E.	*n*
11	**1.023**	0.013										0.013	0.020					0.048	0.040	1.157	0.88	0.04	91
21		**0.420**	0.001	0.059			0.013	0.028	0.001				0.027				0.019			0.568	0.74	0.07	120
22		0.057	**0.005**	0.025				0.001					0.001				0.002			0.091	0.05	0.05	69
23		0.038		**0.193**			0.013	0.013			0.006	0.006	0.013		0.019	0.006	0.019		0.006	0.332	0.58	0.07	59
31	0.001			0.001	**0.011**	0.001	0.004		0.001	0.001	0.002	0.021	0.001		0.001	0.002			0.001	0.048	0.24	0.05	52
32				0.002		**0.026**	0.003	0.002				0.001	0.004						0.001	0.039	0.67	0.11	87
33							**0.001**	0.004					0.001							0.006	0.09	0.07	78
41	0.022	0.110		0.277			0.115	**5.924**	0.626	0.597	0.004	0.111	1.646		0.106	0.013	0.009	0.159	0.022	9.741	0.61	0.06	177
42							0.006	0.162	**0.323**	0.072			0.022		0.013				0.006	0.604	0.53	0.11	79
43				0.018				0.145	0.068	**0.287**			0.052			0.009				0.579	0.50	0.11	55
51		0.004		0.004			0.004	0.013	0.008		**0.076**	0.021	0.055			0.008				0.193	0.39	0.07	46
71	0.001	0.006		0.299	0.007		0.293	0.006	0.302	0.270	2.639	**13.980**	4.069	0.165	2.089	4.379	0.006	0.285	0.978	29.774	0.47	0.05	232
81		0.004		0.010	0.004			0.712	0.009			0.345	**12.864**	0.165	1.291	1.322		0.027	0.178	16.391	0.76	0.04	162
82				0.634	0.002			0.515			0.011	0.285	14.383	**1.537**	11.367	1.601	0.006	0.009	0.630	30.980	0.05	0.02	158
83				0.073		0.070		0.070					1.116		**0.143**	5.001			0.070	6.543	0.02	0.02	102
84													0.056		0.050	**0.162**				0.268	0.61	0.08	87
85		0.057		0.056			0.031	0.029	0.003				0.084		0.003		**0.149**			0.412	0.36	0.18	90
91					0.115			0.028	0.009	0.027		0.009	0.053			0.027		**0.561**	0.132	0.961	0.58	0.11	76
92	0.023			0.010			0.029	0.019			0.019	0.078	0.029	0.010	0.023	0.019		0.029	**0.482**	0.770	0.63	0.06	80
Total	1.070	0.709	0.006	1.661	0.139	0.097	0.512	7.671	1.350	1.254	2.757	14.870	34.496	1.877	15.105	12.549	0.210	1.118	2.546	99.997			
Prod.	0.96	0.59	0.80	0.12	0.08	0.27	0.00	0.77	0.24	0.23	0.03	0.94	0.37	0.82	0.01	0.01	0.71	0.50	0.19				
S.E.	0.02	0.11	0.18	0.05	0.07	0.21	0.00	0.06	0.08	0.09	0.01	0.03	0.03	0.12	0.01	0.00	0.16	0.15	0.06			0.02	
n	86	156	5	117	27	51	47	225	80	54	53	150	370	11	92	190	26	72	88				1900

Midwest, Anderson Level I[a]

	10	20	30	40	50	80	90	Total	Users	S.E.	*n*
10	**0.988**	0.013			0.013	0.006	0.075	1.095	0.90	0.04	86
20		**0.789**	0.025	0.054	0.006	0.073		0.947	0.83	0.03	241
30	0.001	0.002	**0.048**	0.007	0.023	0.009	0.002	0.092	0.52	0.07	216
40	0.022	0.406	0.029	**8.464**	0.115	1.261	0.161	10.458	0.81	0.04	315
50	0.001	0.305	0.304	0.600	**16.569**	9.160	1.268	28.207	0.59	0.05	272
80		0.824	0.102	1.100	0.352	**54.191**	0.924	57.493	0.94	0.01	619
90	0.045	0.010	0.135	0.068	0.106	0.143	**1.201**	1.708	0.70	0.06	151
Total	1.057	2.349	0.643	10.293	17.184	64.843	3.631	100.000			
Prod.	0.93	0.34	0.07	0.82	0.96	0.84	0.33				
S.E.	0.03	0.09	0.04	0.04	0.02	0.02	0.07			0.02	
n	85	272	128	362	200	697	156				1900

Error matrix for Level I and Level II classification: Midwest region. Level II overall accuracy was 38% with a S.E. of 2% and Level I overall accuracy was 82% with a S.E. of 2%. See Table 3 for description of contents.

[a] Shrubland (51) and grassland (71) combined into single Anderson Level I class (50).

Table 5
South-central region, Anderson Level II

	11	21	22	23	31	32	33	41	42	43	51	61	71	81	82	83	84	85	91	92	Total	Users	S.E.	*n*
11	**1.665**			0.021	0.041			0.038	0.021				0.062	0.021			0.041		0.082	0.104	2.096	0.80	0.05	103
21	0.006	**0.332**	0.016	0.073			0.001	0.031	0.026	0.008	0.019	0.006	0.013	0.016		0.006	0.013	0.055	0.006		0.627	0.53	0.06	121
22	0.002	0.076	**0.005**	0.038					0.004				0.002	0.008				0.023			0.158	0.03	0.02	81
23		0.009	0.005	**0.241**	0.005			0.014	0.005	0.009	0.005		0.009	0.023				0.037	0.005	0.006	0.373	0.65	0.05	88
31	0.056			0.032	**0.257**		0.008				0.145		0.349	0.008			0.008		0.008	0.016	0.887	0.29	0.09	82
32				0.006		**0.008**	0.001		0.001				0.002	0.002						0.001	0.021	0.38	0.10	94
33				0.004			**0.184**	0.017	0.082	0.017	0.073	0.004	0.009	0.017			0.004		0.009	0.009	0.429	0.43	0.08	92
41				0.005			0.130	**3.167**	1.365	0.576	0.533		0.978	1.451			0.087	0.096	1.005	0.175	9.568	0.33	0.05	121
42		0.017			0.008		0.195	0.381	**5.554**	1.300	0.610	0.229	0.476	0.278			0.191		0.448		9.677	0.57	0.06	117
43	0.038	0.005					0.076	0.571	1.026	**0.609**	0.190			0.076					0.306	0.076	2.973	0.21	0.05	79
51		0.243			0.245	0.005		0.420	1.850	0.474	**12.326**		5.794	1.192				0.337		0.237	23.123	0.53	0.05	124
61												**0.000**							0.001		0.001	0.00	0.00	85
71			0.223		0.008		0.223		0.683		5.848		**8.687**	5.613			0.714	0.005			22.004	0.40	0.05	118
81		0.209		0.025			0.138	0.289	0.326	0.183	0.480		0.415	**5.840**		0.096	1.156	0.101	0.096	0.766	10.120	0.58	0.05	130
82				0.103				0.095			0.095		0.190	2.932	**2.376**	1.711	1.776	0.096		0.468	9.842	0.24	0.05	124
83				0.049							0.045		0.357	0.895		**0.313**	2.640		0.089	0.110	4.498	0.07	0.03	103
84														0.006		0.001	**0.014**			0.001	0.022	0.61	0.08	81
85		0.002	0.001	0.006						0.001	0.001		0.004	0.008			0.001	**0.038**	0.001	0.005	0.068	0.56	0.05	71
91									0.039	0.275	0.020		0.039	0.039					**1.219**	0.079	1.965	0.62	0.05	94
92	0.034						0.017	0.255	0.017	0.017	0.017		0.017	0.017		0.017	0.017	0.005	0.471	**0.847**	1.544	0.55	0.09	92
Total	1.801	0.893	0.250	0.603	0.564	0.013	0.973	5.329	10.989	3.469	20.407	0.239	17.403	18.442	2.367	2.144	6.662	0.793	3.746	2.900	99.996			
Prod.	0.92	0.37	0.02	0.40	0.46	0.64	0.19	0.59	0.50	0.18	0.60	0.00	0.50	0.32	1.00	0.15	0.00	0.05	0.33	0.29				
S.E.	0.03	0.12	0.02	0.08	0.20	0.24	0.07	0.06	0.05	0.05	0.05	0.00	0.05	0.03	0.00	0.05	0.00	0.02	0.05	0.07			0.02	
n	97	113	10	146	39	37	60	118	168	73	156	7	136	246	25	44	166	88	157	114				2000

South-central region, Anderson Level I[a]

	10	20	30	40	50	80	90	Total	Users	S.E.	*n*
10	**1.665**	0.021	0.041	0.021	0.062	0.062	0.186	2.057	0.81	0.05	101
20	0.006	**0.807**	0.009	0.083	0.048	0.172	0.013	1.139	0.71	0.04	288
30	0.057	0.042	**0.397**	0.113	0.565	0.044	0.042	1.260	0.32	0.07	262
40	0.038	0.263	0.443	**15.004**	2.347	2.485	2.016	22.596	0.66	0.03	330
50	0.002	0.234	0.481	2.635	**33.897**	7.862	0.241	45.352	0.75	0.03	253
80		0.387	0.155	0.856	1.364	**19.894**	1.538	24.193	0.82	0.02	586
90	0.034			0.615	0.093	0.092	**2.569**	3.403	0.76	0.04	180
Total	1.802	1.754	1.526	19.327	38.376	30.611	6.605	100.000			
Prod.	0.92	0.46	0.26	0.78	0.88	0.65	0.39				
S.E.	0.03	0.11	0.07	0.03	0.02	0.03	0.05			0.02	
n	96	271	133	360	297	575	268				2000

Error matrix for Level I and Level II classification: South-central region. Level II overall accuracy was 44% with a S.E. of 2%, and Level I overall accuracy was 74% with a S.E. of 2%. See Table 3 for description of contents.

[a] Shrubland (51) and grassland (71) combined into single Anderson Level I class (50). Orchard (61) is included in a single agricultural class (80).

Table 6
Rocky Mountain region, Anderson Level II

	11	12	21	22	23	31	32	33	41	42	43	51	71	81	82	83	84	85	91	92	Total	Users	S.E.	*n*
11	**1.201**									0.032		0.032	0.081		0.048	0.016			0.016	0.145	1.571	0.76	0.08	100
12		**0.004**				0.027				0.001		0.007	0.010								0.049	0.09	0.10	81
21			**0.205**	0.001	0.022			0.014	0.003	0.003		0.006	0.023		0.010			0.003			0.290	0.71	0.05	118
22	0.002		0.018	**0.022**	0.002			0.001				0.001						0.001			0.047	0.45	0.12	88
23	0.003		0.013	0.010	**0.072**	0.003	0.003	0.003				0.005	0.084	0.003	0.028	0.005					0.232	0.31	0.13	67
31	0.018					**1.187**				0.391	0.089	0.266	0.266		0.018				0.018	0.160	2.413	0.49	0.13	94
32						0.001	**0.022**	0.003		0.001		0.002	0.003	0.001	0.051						0.084	0.26	0.19	98
33								**0.004**	0.008	0.427	0.030	0.042	0.061								0.572	0.01	0.01	97
41						0.025			**0.340**	0.882	0.148	0.382	0.205		0.076	0.025					2.083	0.16	0.05	98
42	0.003					0.001			1.084	**8.971**	1.318	2.498	0.584								14.459	0.62	0.06	140
43									0.022	0.074	**0.029**	0.042	0.019						0.003		0.189	0.15	0.07	59
51		0.001	0.207		0.228	0.846	0.002			1.291		**15.381**	1.226		0.207						19.389	0.79	0.05	117
71	0.002	0.207	0.001		0.358	0.225			0.355	0.382	0.071	4.881	**22.802**	0.360	3.975	0.720			0.018		34.996	0.65	0.05	159
81			0.051		0.088		0.050		0.005	0.053		0.405	1.026	**1.548**	1.259	0.050					4.580	0.34	0.06	108
82					0.003								0.323	0.989	**5.969**	0.313					7.597	0.79	0.05	107
83					0.072							0.287	1.021	0.072	2.639	**2.639**					6.730	0.39	0.06	99
84												0.267	0.646	0.530	1.712	0.530	**0.000**				3.685	0.00	0.00	94
85			0.019		0.024							0.009	0.010	0.002	0.006			**0.011**			0.081	0.14	0.04	85
91	0.009					0.007			0.009	0.019	0.005	0.037	0.046	0.009	0.014				**0.007**	0.002	0.164	0.04	0.04	71
92	0.031	0.010				0.010			0.002	0.041	0.010	0.158	0.112	0.031	0.245	0.031			0.010	**0.102**	0.793	0.13	0.07	80
Total	1.269	0.222	0.514	0.033	0.869	2.332	0.077	0.025	1.873	12.568	2.339	24.708	28.548	3.545	16.257	4.329	0.000	0.15	0.072	0.409	100.00			
Prod.	0.95	0.02	0.40	0.66	0.08	0.51	0.28	0.15	0.18	0.71	0.01	0.62	0.80	0.44	0.37	0.61	0.00	0.74	0.10	0.25				
S.E.	0.02	0.03	0.18	0.25	0.05	0.15	0.22	0.14	0.07	0.05	0.01	0.04	0.03	0.08	0.04	0.09	0.00	0.17	0.10	0.17			0.02	
n	93	10	143	45	84	106	65	19	45	242	47	281	306	78	279	67	0	13	8	29				1960

Rocky Mountain region, Anderson Level I[a]

	10	20	30	40	50	80	90	Total	Users	S.E.	*n*
10	**1.193**		0.027	0.033	0.129	0.064	0.161	1.607	0.74	0.08	178
20	0.005	**0.370**	0.023	0.006	0.116	0.049		0.569	0.65	0.10	273
30	0.018		**1.184**	0.942	0.600	0.070	0.178	2.992	0.40	0.11	281
40	0.003		0.025	**13.246**	3.052	0.101	0.003	16.430	0.81	0.04	297
50	0.208	0.583	0.434	2.359	**45.950**	5.264	0.018	54.816	0.84	0.03	289
80		0.097	0.050	0.050	2.496	**19.965**		22.658	0.88	0.02	497
90	0.042		0.017	0.084	0.349	0.304	**0.132**	0.928	0.14	0.08	145
Total	1.468	1.051	1.761	16.719	52.692	25.817	0.491	100.000			
Prod.	0.81	0.35	0.67	0.79	0.87	0.77	0.27				
S.E.	0.12	0.16	0.14	0.05	0.01	0.04	0.16			0.02	
n	100	268	184	343	553	474	38				1960

Error matrix for Level I and Level II classification: Rocky Mountain region. Level II overall accuracy was 61% with a S.E. of 2%, and Level I overall accuracy was 82% with a S.E. of 2%. See Table 3 for description of contents.

[a] Shrubland (51) and grassland (71) combined into single Anderson Level I class (50).

Table 7

Southwest region, Anderson Level II

	11	12	21	22	23	31	32	33	41	42	43	51	61	71	81	82	83	84	85	91	92	Total	Users	S.E.	*n*
11	**0.554**		0.008		0.016	0.024			0.025	0.008		0.008	0.008	0.024	0.008	0.039	0.024				0.025	0.771	0.72	0.11	101
12		**0.000**				0.002																0.002	0.00	0.00	66
21			**0.509**	0.043	0.063	0.007					0.007	0.015		0.037	0.007	0.007	0.001				0.001	0.697	0.73	0.04	126
22			0.016	**0.017**	0.014												0.002					0.049	0.34	0.14	71
23	0.004		0.026	0.008	**0.182**	0.004		0.004				0.018		0.018	0.004	0.004	0.011					0.283	0.65	0.09	95
31			0.050		0.050	**1.111**						2.133		1.528							0.102	4.974	0.22	0.07	98
32	0.001			0.001	0.020	0.001	**0.035**	0.012	0.108			0.006	0.001	0.024	0.001	0.001			0.004			0.215	0.16	0.14	96
33			0.001					**0.005**		0.004	0.001	0.001		0.090								0.014	0.33	0.13	76
41			0.015					0.016	**0.081**	0.066	0.074	0.155		0.033								0.440	0.18	0.07	51
42	0.008		0.197			0.524			1.803	**8.899**	1.346	4.052		0.524								17.353	0.51	0.06	180
43						0.015			0.120	0.280	**0.537**	0.166	0.023	0.090	0.015							1.246	0.43	0.12	63
51			0.009	0.001	0.005	3.372	0.001	0.003	0.106	1.960	0.600	**49.082**	0.013	1.230		0.013	0.004				0.575	56.974	0.86	0.03	241
61			0.013		0.026	0.013		0.008				0.026	**0.476**	0.013	0.052	0.459	0.229					1.315	0.36	0.08	104
71			0.023		0.318	0.846		0.121	0.330	0.706	0.523	1.622	0.209	**6.198**		0.053	0.021		0.105			11.075	0.56	0.06	144
81	0.032		0.072		0.001			0.021	0.021	0.021		0.107	0.044	0.473	**0.865**	0.331	0.120				0.021	2.129	0.41	0.06	123
82			0.037	0.012					0.012			0.050	0.086	0.077	0.197	**0.438**	0.386					1.295	0.34	0.06	102
83	0.020		0.010					0.020	0.010	0.010		0.020	0.020	0.051	0.041	0.124	**0.629**					0.955	0.66	0.08	95
84												0.001				0.001		**0.000**				0.002	0.00	0.00	40
85			0.018	0.002	0.005		0.001		0.001			0.008	0.001	0.003			0.001		**0.020**		0.001	0.061	0.34	0.12	68
91						0.003			0.007			0.013	0.003	0.002						**0.004**		0.032	0.12	0.05	71
92	0.009		0.001		0.003	0.017		0.004		0.003		0.010		0.009	0.007	0.003	0.017				**0.030**	0.113	0.27	0.09	80
Total	0.628	0.000	1.005	0.084	0.703	5.939	0.037	0.214	2.624	11.957	3.088	57.493	0.884	10.336	1.197	1.473	1.445	0.000	0.129	0.004	0.755	99.995			
Prod.	0.88	0.00	0.51	0.20	0.26	0.19	0.95	0.02	0.03	0.74	0.17	0.85	0.54	0.60	0.72	0.30	0.44	0.00	0.16	1.00	0.04				
S.E.	0.07	0.00	0.11	0.06	0.11	0.06	0.02	0.02	0.01	0.07	0.07	0.02	0.11	0.08	0.06	0.06	0.07		0.14	0.00	0.04			0.02	
n	85	0	167	38	139	151	32	54	74	161	72	376	70	230	90	125	151	0	34	9	33				2091

Southwest region, Anderson Level I[a]

	10	20	30	40	50	80	90	Total	Users	S.E.	*n*
10	**0.554**	0.024	0.026	0.032	0.032	0.071	0.025	0.763	0.73	0.11	162
20	0.004	**0.871**	0.011	0.007	0.081	0.035	0.001	1.011	0.86	0.04	290
30		0.119	**1.161**	0.007	3.690	0.007	0.102	5.087	0.23	0.07	259
40	0.008	0.030	0.389	**15.065**	3.399	0.023		18.914	0.80	0.04	299
50		0.051	1.777	3.269	**62.238**	0.371	0.575	68.281	0.91	0.02	398
80	0.031	0.155	0.043	0.066	0.749	**4.732**	0.022	5.798	0.82	0.04	531
90	0.009	0.004	0.024	0.11	0.029	0.030	**0.039**	0.146	0.27	0.08	152
Total	0.605	1.254	3.430	18.458	70.220	5.270	0.764	100.000			
Prod.	0.92	0.70	0.34	0.82	0.89	0.90	0.05				
S.E.	0.05	0.08	0.11	0.05	0.02	0.04	0.04			0.02	
n	83	328	222	328	591	487	52				2091

Error matrix for Level I and Level II classification: Southwest region. Level II overall accuracy was 70% with a S.E. of 2%, and Level I overall accuracy was 85% with a S.E. of 2%. See Table 3 for description of contents.

[a] Shrubland (51) and grassland (71) combined into single Anderson Level I class (50). Orchard (61) is included in single agricultural class (80).

Table 8
Pacific Northwest, Anderson Level II

	11	12	21	22	23	31	32	33	41	42	43	51	61	71	81	82	83	84	85	91	92	Total	Users	S.E.	*n*
11	**2.866**				0.032	0.065			0.032					0.065							0.032	3.092	0.93	0.03	99
12		**0.118**			0.001	0.009																0.129	0.92	0.03	97
21	0.004		**0.390**	0.012	0.022	0.006	0.006		0.026	0.012	0.004	0.006	0.006	0.030	0.030			0.006	0.030	0.006		0.597	0.65	0.07	114
22			0.001	**0.000**	0.001														0.001			0.002	0.06	0.04	76
23			0.019	0.004	**0.193**	0.006	0.004	0.004	0.004	0.012		0.004		0.016	0.014			0.012		0.004	0.004	0.301	0.64	0.07	77
31	0.094	0.142				**1.245**				0.234		0.188		0.648				0.023			0.047	2.621	0.47	0.12	91
32					0.001		**0.013**															0.015	0.85	0.06	114
33			0.017			0.050	0.017	**1.749**		0.132	0.017	0.050		0.050	0.017			0.017				2.113	0.83	0.06	84
41			0.032		0.010			0.449	**0.466**	0.537	0.252	0.056		0.112	0.559			0.028		0.002	0.028	2.531	0.18	0.05	88
42	0.002		0.006		0.397	0.300		2.196	0.448	**29.416**	1.264	1.523		0.834				0.002		0.064		36.452	0.81	0.04	190
43	0.029		0.035		0.057	0.029		0.201	0.344	0.488	**0.282**	0.086	0.029		0.029				0.029		0.002	1.638	0.17	0.05	60
51	0.302		0.001		0.009	0.600		0.006	0.300	0.685	0.300	**14.175**		10.796	0.046		0.610	0.663			0.085	28.579	0.50	0.06	148
61			0.002		0.008			0.008	0.002	0.004	0.002	0.004	**0.057**	0.004	0.065	0.006	0.016	0.004	0.004		0.002	0.187	0.30	0.08	95
71		0.085	0.085					0.843	0.002	0.595		0.542		**5.469**	0.087	0.300		0.441			0.097	8.547	0.64	0.06	124
81	0.042				0.046	0.042			0.125	0.167	0.042	0.125	0.042	0.393	**1.980**	0.044	0.250	0.674	0.004		0.210	4.186	0.47	0.05	123
82					0.016					0.016		0.032		0.016	0.207	**0.064**	0.032	1.186				1.567	0.04	0.02	97
83					0.041	0.037		0.426		0.037		0.037		0.037	1.207	0.148	**1.559**	0.635				4.164	0.37	0.06	99
84														0.026	0.078		0.026	**2.426**				2.556	0.95	0.02	101
85			0.001		0.001							0.001		0.002	0.005				**0.015**			0.026	0.57	0.09	78
91	0.004		0.002			0.004		0.002	0.010	0.011	0.004	0.006		0.016					0.002	**0.021**	0.033	0.113	0.19	0.07	57
92	0.060					0.012				0.012		0.074		0.043	0.031	0.006		0.019		0.006	**0.318**	0.583	0.55	0.10	89
Total	3.402	0.345	0.589	0.016	0.836	2.404	0.041	5.884	1.759	32.361	2.166	16.909	0.134	18.558	4.353	0.568	2.493	6.136	0.084	0.103	0.858	100.000			
Prod.	0.84	0.34	0.66	0.01	0.23	0.52	0.31	0.30	0.26	0.91	0.13	0.84	0.43	0.29	0.45	0.11	0.62	0.39	0.17	0.20	0.37				
S.E.	0.08	0.14	0.14	0.01	0.12	0.15	0.17	0.09	0.09	0.02	0.06	0.04	0.18	0.04	0.06	0.08	0.11	0.05	0.08	0.11	0.12			0.02	
n	108	97	116	9	133	72	108	116	56	233	36	123	32	207	174	16	63	231	66	19	86				2101

Pacific Northwest, Anderson Level I[a]

	10	20	30	40	50	80	90	Total	Users	S.E.	*n*
10	**2.982**	0.034	0.074	0.032	0.065		0.032	3.219	0.93	0.03	195
20	0.004	**0.623**	0.026	0.048	0.057	0.083	0.014	0.856	0.73	0.05	264
30	0.236	0.017	**2.704**	0.351	0.912	0.057	0.047	4.324	0.63	0.08	278
40	0.002	0.507	2.977	**34.226**	2.586	0.679	0.068	41.044	0.83	0.03	352
50	0.410	0.097	1.326	1.859	**31.162**	2.145	0.182	37.182	0.84	0.03	281
80	0.042	0.072	0.469	0.270	0.513	**11.200**	0.133	12.699	0.88	0.03	591
90	0.060	0.002	0.016	0.032	0.121	0.058	**0.387**	0.676	0.57	0.08	140
Total	3.736	1.352	7.593	36.818	35.415	14.221	0.865	100.000			
Prod.	0.80	0.46	0.36	0.93	0.88	0.79	0.45				
S.E.	0.08	0.15	0.08	0.02	0.03	0.05	0.12			0.02	
n	202	254	279	336	323	600	107				2101

Error matrix for Level I and Level II classification: Pacific Northwest region. Level II overall accuracy was 63% with a S.E. of 2%, and Level I overall accuracy was 83% with a S.E. of 2%. See Table 3 for description of contents.

[a] Shrubland (51) and grassland (71) combined into single Anderson Level I class (50). Orchard (61) is included in single agricultural class (80).

A high proportion of the error in the Level II matrices was confusion between classes that aggregate to the same Level I group. Thus, despite the low and variable (among regions) Level II overall accuracy scores, Level I overall accuracy improved to greater than 80% for five of the six mapping regions, ranging from 82% to 85%. Except for water, Level I user's accuracies were highest for agriculture, ranging from 82% to 94% across the six western mapping regions. Level I user's accuracies for forest were between 80% and 83% for the five western mapping regions where Level I overall accuracy equaled or exceeded 80%. Level I user's accuracies for urban were variable ranging from 65% to 92% across the six western mapping regions. The shrubland/grassland user's accuracies were high in the three western mapping regions where the class was areally dominant (84% Rocky Mountains; 91% Southwest; 84% Pacific Northwest). Level I overall accuracy for the South-central region was lower at 74% (Table 5), primarily because of confusion between "semi-natural" grassland and hay/pasture (6%).

The general absence of zero values for off-diagonal cell entries of the Level I error matrices indicated that although a high proportion of error was attributable to classes that aggregate to a common Level I group (within-Level I class), there was still considerable confusion between Level I groups. For the regions reported here, within- and between-Level I error rates were about equal at 19% and 17% (Great Lakes), 26% and 18% (Midwest), 30% and 26% (South-central), 21% and 18% (Rocky Mountains), 15% and 15% (Southwest), and 20% and 17% (Pacific Northwest), respectively. For the Midwest, within- and between-Level I errors were 26% and 18% based on Level II overall accuracy of 56% realized by treating row crops (82), small grains (83), and fallow (84) as a single class. The pattern of within- and between-Level I error across the six regions suggests that the likelihood of misclassification between Level II components comprising a Level I class was about equal to the likelihood of misclassification among the Level I classes themselves. The primary errors between Level I classes were confusion between forest (40) and shrubland/

Table 9
Estimated accuracy (%) for various definitions of agreement for the Great Lakes region

Class	User's accuracy				Producer's accuracy			
Level II	CenterP	Center	Mode	Homogeneous	CenterP	Center	Mode	Homogeneous
11	94	94	96	100	86	88	91	94
21	73	77	85	100	29	34	40	67
22	11	21	24	33	72	93	89	89
23	63	69	49	79	22	26	24	61
31	45	51	59	27	1	8	6	100
32	23	32	33	43	49	56	55	89
33	33	41	56	64	3	4	5	4
41	46	56	61	80	71	78	86	90
42	36	54	63	80	36	51	52	44
43	51	63	74	80	20	27	28	15
51	20	39	52	38	1	1	1	0
71	3	9	33	20	1	3	13	3
81	6	13	16	16	53	70	81	91
82	84	89	90	100	74	77	78	88
83	10	61	63	60	5	28	50	35
85	36	51	57	75	10	16	15	18
91	34	39	42	41	40	49	50	75
92	25	37	35	54	28	41	38	53
All	53	60	64	79				
Level I								
10	94	94	96	100	86	90	91	93
20	81	85	92	97	39	55	62	64
30	34	44	49	49	6	11	11	16
40	63	76	81	88	70	77	82	87
50	15	25	54	24	2	5	9	1
80	80	89	89	98	93	95	97	99
90	40	50	53	53	48	66	67	80
All	70	80	83	91				

The definitions of agreement are described in Section 2.2. The "Mode" definition of agreement is the same as the one used to derive the error matrices in Tables 3–8.

The row denoted "All" has overall accuracy for the region based on the column heading's definition of agreement. The sample size for "CenterP", "Center", and "Mode" agreement definitions was 1634 at Level I and II, and the sample size for the "Homogeneous" definition of agreement was 492 at Level II and 760 at Level I.

grassland (50) or shrubland/grassland and agriculture (80) (Tables 3–8).

3.2. Comparing accuracy for different definitions of agreement

The accuracy estimates reflect the expected progression of lower accuracy for the CenterP definition of agreement to higher accuracy for the homogeneous subset (Tables 9–11). The increase in accuracy of the Center estimates relative to the CenterP estimates quantifies the effect of incorporating the alternate reference land-cover label (i.e., reference land-cover ambiguity) into the agreement definition. The accuracy estimates based on the Center and Mode definitions of agreement were generally similar (save a few exceptions), indicating that allowing for spatial registration error (the Mode definition) did not markedly change the results. The higher accuracies of the homogeneous subset must be interpreted with the recognition that they do not represent the entire map (i.e., the population sampled). Comparing the homogeneous subset with the other agreement definitions provides useful quantitative information on the likely spatial distribution of errors across the entire map. The overall sharp increases in accuracy for the homogeneous subset as compared to the other three agreement definitions are a general indication that a substantial fraction of the disagreement between map and reference labels occurs at the edges between mapped land-cover classes (Smith et al., 2002, 2003). Similarly sharp increases in accuracy occur between the CenterP and Center agreement definitions. The results in Tables 9–11 suggest that ambiguity in reference land-cover labels and land-cover heterogeneity in the map account for a substantial portion of the misclassification reported in Tables 3–8.

3.3. Comparison of eastern and western mapping regions

The Level II and I overall accuracies for the six western mapping regions were in general agreement with those reported for the four eastern mapping regions (Stehman et al., 2003). Level II overall accuracies for the eastern U.S. regions ranged from 43% to 66%, but the Level I overall

Table 10
Estimated accuracy (%) for various definitions of agreement for the Rocky Mountain region

Class	User's accuracy				Producer's accuracy			
Level II	CenterP	Center	Mode	Homogeneous	CenterP	Center	Mode	Homogeneous
11	72	74	76	82	72	94	95	97
12	4	8	9	0	1	2	2	0
21	59	68	71	68	34	38	40	18
22	24	42	45	68	38	61	66	75
23	29	30	31	50	6	9	8	12
31	25	38	49	35	23	43	51	29
32	60	62	26	76	28	28	28	88
33	1	1	1	1	15	15	15	33
41	8	13	16	16	9	17	18	7
42	53	61	62	63	58	72	71	87
43	10	15	15	0	1	2	1	0
51	62	77	79	85	41	62	62	54
71	34	67	65	69	56	80	80	88
81	21	36	34	34	28	47	44	38
82	71	79	79	80	31	36	37	44
83	17	38	39	40	34	59	61	79
84	0	0	0	0	0	0	0	0
85	9	9	14	10	47	55	74	100
91	3	3	4	0	6	10	10	0
92	2	11	13	17	6	25	25	14
All	41	60	61	65				
Level I								
10	69	71	74	81	64	80	81	97
20	62	64	65	86	21	35	35	43
30	22	33	40	28	24	56	67	48
40	70	77	81	82	65	79	79	84
50	74	84	84	87	79	86	87	90
80	77	88	88	90	73	77	77	82
90	8	11	14	15	20	27	27	14
All	72	81	82	82				

Agreement definitions are the same as in Table 9. The sample size for "CenterP", "Center", and "Mode" agreement definitions was 1960 at Level I and II, and the sample size for the "Homogeneous" definition of agreement was 857 at Level II and 1193 at Level I.

Table 11
Estimated accuracy (%) for various definitions of agreement for the Southwest region

Class	User's accuracy				Producer's accuracy			
Level II	CenterP	Center	Mode	Homogeneous	CenterP	Center	Mode	Homogeneous
11	67	71	72	87	89	89	88	97
12	0	0	0	0	0	0	0	0
21	52	67	73	76	36	50	51	60
22	19	26	34	10	7	17	20	3
23	43	52	65	88	18	26	26	30
31	19	29	22	29	10	24	19	21
32	21	31	16	44	95	97	95	100
33	32	39	33	40	3	3	2	1
41	12	21	18	33	3	6	3	1
42	46	53	51	56	63	76	74	80
43	20	41	43	33	9	19	17	5
51	68	86	86	86	79	86	85	91
61	35	36	36	39	51	52	54	69
71	28	50	56	71	21	51	60	48
81	14	38	41	39	39	69	72	69
82	26	33	34	39	22	28	30	32
83	43	62	66	71	25	43	44	45
84	0	0	0	0	0	0	0	0
85	26	30	34	67	13	16	16	100
91	0	9	12	11	0	100	100	100
92	21	24	27	22	4	5	4	1
All	53	69	70	75				
Level I								
10	67	71	73	87	89	93	92	97
20	74	78	86	91	41	71	70	78
30	20	30	23	30	11	40	34	31
40	64	80	80	87	66	83	82	91
50	77	92	91	94	83	89	89	93
80	76	81	82	87	87	88	90	96
90	18	23	27	22	4	6	5	2
All	72	85	85	90				

Agreement definitions are the same as in Table 9. The sample size for "CenterP", "Center", and "Mode" agreement definitions was 2091 at Level I and II, and the sample size for the "Homogeneous" definition of agreement was 743 at Level II and 1038 at Level I.

accuracies were generally above 80%. Three class-specific differences in error patterns between eastern and western mapping regions were identified. First, eastern mapping region Level I user's accuracies were distinctly higher for urban, agriculture, and forest relative to user's accuracies for barren and wetland (Stehman et al., 2003), a disparity not evident in the western regions. Second, the eastern mapping regions also showed small but distinct confusion between wetland and upland forest classes, and Level I overall accuracies for the eastern U.S. could have been improved marginally by labeling wetland forest (92) as upland forest (40). This pattern also was not evident in the western mapping regions. Third, four of the six western mapping regions had distinctly low user's accuracies for water (< 90%), and these reduced user's accuracies were not solely attributable to the presence of the perennial ice/snow class. The Great Lakes mapping region was more similar to the eastern U.S. regions in regard to the three class-specific differences in error between east and west. The Great Lakes mapping region had: (1) distinctly higher Level I user's accuracies for urban, agriculture, and forest; (2) the expected high user's accuracy for water (96%); and (3) small but distinct errors between upland and wetland forest (Table 3).

4. Discussion

Anderson et al. (1976, p. 5) proposed a nominal standard of 85% as a threshold of acceptable minimum accuracy for both Level I and Level II classifications. Eight of the 10 regional overall accuracies for the 1992 NLCD reported here and in Stehman et al. (2003) come close to meeting this standard at Level I, but not at Level II. Congalton and Green (1993) reviewed eight factors that can affect error matrix results, and these factors provide a useful framework for further interpretation of the results presented in Tables 3–11. The factors were: (1) land-cover change between the time of reference data acquisition and satellite data acquisition, (2) reference sample location error, (3) reference label data entry error, (4) reference label photointerpretation error, (5) inconsistent labeling of reference data due to land-cover heterogeneity surrounding the sample location, (6) differences in map and reference data registration, (7) map delineation error, and (8) map classification error. The first factor relates to errors arising from change over time, factors (2)–(6) relate to errors in the reference data themselves, and Congalton and Green (1993) identified the last two factors as genuine classification error.

Reference and satellite data acquisition dates were known for about half (6052) of the 11,975 samples. Differences in acquisition dates (factor 1) did not strongly affect agreement results. The mean absolute difference between photo and satellite acquisition dates for pixel groups defined by agreement (correctly classified pixels) and disagreement (misclassified pixels) was statistically different ($\alpha = 0.05$) for seven of the 21 NLCD classes. However, for four of the seven classes the mean differences were in the wrong direction (Table 12). That is, the mean differences between photo and satellite acquisition dates were smaller for the disagreement group, the opposite relationship typically associated with the effect of temporal differences in acquisition dates on accuracy. Only perennial snow/ice (12), high-density urban (12), and shrubland (51) had significantly different means between agreement and disagreement groups and higher means for the disagreement group.

Table 12
Difference in satellite and photo acquisition dates

Class	Disagreement		Agreement		p-Value
	n	Mean	n	Mean	
11	58	1.82	233	2.45	**0.01**
12	163	3.46	128	2.28	**< 0.01**
21	105	2.15	197	2.03	0.44
22	198	3.63	101	2.63	**< 0.01**
23	124	2.36	171	2.32	0.82
31	164	1.99	124	2.54	**< 0.01**
32	126	1.70	173	1.59	0.33
33	175	3.74	110	3.65	0.79
41	200	2.40	86	2.29	0.67
42	103	2.28	184	2.79	**0.04**
43	179	2.19	114	2.25	0.73
51	89	2.66	205	2.20	**0.02**
61	131	2.31	67	2.03	0.13
71	104	2.18	185	2.54	0.09
81	170	2.02	124	2.07	0.74
82	179	2.07	112	1.74	0.08
83	151	2.62	142	2.35	0.20
84	157	2.18	134	2.66	**0.01**
85	174	2.49	126	2.33	0.36
91	219	2.59	76	2.78	0.40
92	192	1.96	100	1.87	0.49

Comparison of temporal difference in satellite and reference data acquisition dates for agreement (correctly classified) and disagreement (misclassified) groups.
The "Mean" column is the average absolute difference (in years) between the satellite and reference data acquisition dates based on a sample size of n pixels.
The p-value evaluates the null hypothesis that the mean absolute difference between satellite and reference data acquisition dates is the same for the populations of pixels in the agreement and disagreement groups (two-sample t-test).

Moreover, the difference between the means of the agreement and disagreement groups for these three classes was only about 6–14 months.

Overall, differences between photo and satellite acquisition dates do not appear to have had a strong effect on accuracy results, which is consistent with similar results for NLCD's mid-Atlantic mapping region (Roth et al., 1999). This was not a surprising result. The likelihood of land-cover change between photo and satellite acquisition dates affecting accuracy estimates is dependent on the geographic extent of the mapping effort. Land-cover change is regarded as rare, and that rarity is amplified as the size of the mapping region increases. Differences in photo and satellite acquisition dates can potentially have a significant impact on accuracy results for localized mapping efforts such as the Rogue River National Forest study (e.g., Congalton & Green, 1993), but the effect of such differences on estimation of thematic accuracy of single-date land-cover maps should decline as the geographic scope of the mapping effort increases.

Our analyses and response designs permitted flexibility in defining agreement, and the range in agreement results quantifies effects of reference data errors (factors (2)–(6)) on map accuracy estimates. For example, opting for the modal-based definition of agreement over the center-based definition of agreement represents one approach to account for reference sample location errors (factor (2)) and reference labeling inconsistencies due to land-cover heterogeneity in the vicinity of the sample pixel (factor (5)). NLCD thematic accuracy estimates also can be produced for the subset of the map representing homogeneous 3×3 pixel neighborhoods by using the corresponding subset of the sample data. Overall Level II and Level I accuracies for the homogeneous subset were about 15% and 8% higher than estimates derived from the full sample, with corresponding per-class increases in user's and producer's accuracies. These results are consistent with those documented for the 3×3 homogeneous subset for the eastern U.S. (Stehman et al., 2003). The higher rate of misclassification of pixels at the boundaries between land-cover classes could be attributed to genuine misclassification due to mixed pixel effects (e.g., map delineation error, factor (7)), reference locational or labeling errors (factors (2) and (5)), or both.

Some fraction of the classification error reported here is likely attributable to errors in the reference data. The framework used by Congalton and Green (1993) is useful for gaining insight about errors in the reference data, but it does not permit quantitative estimates of separate components of error because many of the factors are confounded. Auxiliary information (e.g., a sub-sample of ground-based reference data) is required to estimate the error in reference label assignment (Stehman, 1996). Auxiliary reference data collection was not considered for the NLCD accuracy assessment because resources were not available for such an effort. Accounting for reference data errors can, in principle, result in a decrease in classification accuracy

(Stehman, 1996), and the assumption that correcting reference data errors would result in improved agreement (e.g., Congalton & Green, 1993, p. 643) should be treated with caution. The results presented in Tables 9–11 provide quantitative insight into possible effects of reference label errors on reported map accuracy, but they do not guarantee that improved reference data (e.g., auxiliary ground visits) would result in higher map accuracies.

Having now completed the assessment for the entire U.S., we have the advantage of hindsight to evaluate where the accuracy assessment could have been improved. Because the NLCD map was completed sequentially by mapping regions, the accuracy assessment also proceeded sequentially. The sequential approach afforded the opportunity to improve the methodology for the later regions completed, but also resulted in some regional variability in the sampling and response designs implemented.

The response design component of the NLCD accuracy assessment encountered several difficulties common to most assessments targeting very large regions. Differences in the frequency of assignment of alternate reference land-cover labels varied considerably among the reference data interpretation teams, a phenomenon found in other studies (McGwire, 1992; Khorram et al., 1999). Multiple interpretation teams are a likely reality of large-area accuracy assessments. Multiple interpretation teams, operating mostly independently, yielded useful ideas on reference data collection that a top–down inflexible approach would not have revealed. Coordinating all groups would have been costly, time-consuming, and nearly impractical even if all groups were actively collecting reference data concurrently. Re-interpreting regions already completed to impose greater uniformity in use of alternate labels was not fiscally possible.

The difficulties of implementing a completely uniform, nationwide response design protocol are symptomatic of having to initiate accuracy assessments without full knowledge of the ultimate resources that will be available. As Scepan (1999) noted, accuracy assessments are often planned wholly or in part as an afterthought to the mapping effort itself, and the NLCD accuracy assessment is not different in this regard. Motivation and subsequent funding for a national accuracy assessment of the 1992 NLCD occurred after mapping for several of the regions was already underway. Usually it is not possible to expend significant resources in planning the assessment (e.g., there is neither time nor funding to evaluate a multitude of sampling design options or to compare various response design protocols). Consequently, assessments characteristically fall short of the desired rigor. Ideally, accuracy assessment would be included in the planning stages of the mapping effort, and concurrent accuracy assessment planning would logically focus on a priori design evaluation. The design evaluation would consider both statistical rigor and practical constraints (Stehman, 2001) to decide sample size, reference medium, and level of effort.

Despite the difficulties of planning and implementing such a large-scale assessment, the NLCD assessment has several strengths. The sampling design was constructed to adhere to rigorous standards of probability sampling. Although the response design could have been improved by imposing greater regional uniformity on assignment of alternate reference labels, the protocol allows for calculating accuracy for various definitions of agreement reflecting different sensitivity to spatial mis-registration and land-cover class ambiguity. This strength points out the reality that a single interpretation of accuracy for a map may be an oversimplification, given the many difficulties associated with accuracy assessment. It would be convenient to produce a single number that could be cited as “the answer” to, “What is the accuracy of the map?” But as demonstrated in the NLCD assessment, reporting a range of accuracy estimates based on different definitions of agreement provides users with a richer base from which to inform their decisions.

An additional strength of the NLCD accuracy assessment effort was the estimation of standard errors, because estimates of precision are not reported routinely as part of thematic accuracy assessments. The reported standard errors for user’s and producer’s accuracy were affected by three factors: (1) realized (not target) sample size for each class (see row and column margins in Tables 3–8), (2) the estimated accuracy, and (3) the distribution of sample pixels among the clusters. Standard errors increased as the realized sample size decreased, as the estimated accuracy approached 0.50, and as the degree of clustering of sample pixels increased. At Level II, classes such as evergreen forest (42), hay/pasture (81) and cropland (82) tended to have lower standard errors because they often had higher estimated accuracies, sample pixels distributed across several PSUs, and realized sample sizes that exceeded the target of 100. Conversely, rarer classes that had accuracies closer to 0.50, sample pixels clustered into fewer PSUs, and large deviations between realized and target sample sizes had higher standard errors (e.g., the urban [21, 22, 23] and barren [31, 32, 33] classes). Overall, Level II standard errors were more variable, ranging from 0.01 to 0.22 (across all regions), than Level I standard errors. Level I standard errors were generally less than 0.05 for all classes across all regions. The overall lower standard errors at Level I result from a larger sample size and an improved distribution of sample pixels across PSUs that was conferred by combining classes.

Although difficulties were encountered in undertaking a national accuracy assessment of the 1992 NLCD, our results provide useful information for future national land-cover mapping and accompanying thematic accuracy assessments. Perhaps the most significant finding of the 1992 NLCD accuracy assessment was the high degree of confusion between Level I classes. Class ambiguity increases as the thematic detail of classification hierarchies increase (e.g., Level I to Level II), and this increasing ambiguity results in lower accuracy. Instead of judging map quality based on pre-defined, nominal accuracy thresholds applied uniformly to all levels of a classification hierarchy, it is more useful to adjust thematic accuracy goals according to the degree of classification detail. A useful goal for the 2001 NLCD mapping effort (http://www.mrlc.gov/) is to focus on improved discrimination between Level I classes. Higher Level I per-class accuracies for 2001 NLCD would include improved classification of urban across all 10 mapping regions, improved discrimination between agriculture and forest in the eastern U.S., and improved discrimination between shrubland/grassland, forest and agriculture in the western U.S.

Acknowledgements

The U.S. Environmental Protection Agency (EPA), through its Office of Research and Development (ORD), funded and performed the research described. This paper has been subjected to the EPA's peer and administrative review and has been approved for publication. The authors thank three anonymous reviewers for their helpful comments on an earlier draft of the paper. The high quality of reference data used in these analysis were provided by S. Khorram of North Carolina State University's Center for Earth Observation and L. Tinney of Lockheed Martin Corporation through EPA contracts 68-C5-0065 and 68-D-00-267, and T. Slonecker (US EPA) and G. Manuel of U.S. Geological Survey through Interagency Agency Agreements DW14939330-01-0 and DW14939339-01-0.

References

Anderson, J. F., Hardy, E. E., Roach, J. T., & Witmer, R. E. (1976). A land use and land cover classification system for use with remote sensor data. U.S. Geological Survey Professional Paper 964, U.S. Geological Survey, Washington, DC, 28 pp.

Congalton, R. G., & Green, K. (1993). A practical look at the sources of confusion in error matrix generation. *Photogrammetric Engineering and Remote Sensing*, *59*, 641–644.

Dobson, J. E., Bright, E. A., Ferguson, R. L., Field, D. W., Wood, L. L., Haddad, K. D., Iredale, H., Jensen, J. R., Klemas, V. V., Orth, R. J., & Thomas, J. P. (1995). *NOAA Coastal Change Analysis Program (C-CAP): Guidance for regional implementation, NOAA Technical Report NMFS 123*. Seattle, WA: U.S. Department of Commerce.

Goward, S. N. (1989). Landsat 1989: Remote sensing at the crossroads. *Remote Sensing of Environment*, *28*, 3–4.

Hammond, T. O., & Verbyla, D. L. (1996). Optimistic bias in classification accuracy assessment. *International Journal of Remote Sensing*, *17*, 1261–1266.

Khorram, S., Biging, G. S., Chrisman, N. R., Colby, D. R., Congalton, R. G., Dobson, J. E., Ferguson, R. L., Goodchild, M. F., Jensen, J. R., & Mace, T. H. (1999). *Accuracy assessment of remote sensing-derived change detection*. Bethesda, MD, USA: American Society for Photogrammetry and Remote Sensing (ASPRS), Monograph, 64 pp.

Loveland, T. R., & Shaw, D. M. (1996). Multiresolution land characterization: Building collaborative partnerships, Gap Analysis: A Landscape Approach to Biodiversity Planning. In J. M. Scott, T. Tear, F. Davis (Eds.), *Proceedings of the ASPRS/GAP Symposium, Charlotte, North Carolina* (pp. 83–89). Moscow, ID: National Biological Service.

McGwire, K. C. (1992). Analyst variability in labeling of unsupervised classifications. *Photogrammetric Engineering and Remote Sensing*, *58*, 1673–1677.

Roth, N., Strebel, D., Krebs, T., Kou, J. (1999). Accuracy assessment of land cover in Region III. Draft report. Prepared by Versar Inc., Columbia, MD, for the U.S. Environmental Protection Agency.

Scepan, J. (1999). Thematic validation of high-resolution global land-cover data sets. *Photogrammetric Engineering and Remote Sensing*, *65*, 1051–1060.

Scott, J. M., Davis, F., Csuti, B., Noss, R., Butterfield, B., Groves, C., Caicco, S., D'Erchia, F., Edwards, T. C., Ulliman, J., & Wright, R. G. (1993). GAP analysis: A geographic approach to the protection of biological diversity. *Wildlife Monographs No. 123, Supplement to the Journal of Wildlife Management*, *57*(1).

Scott, J. M., & Jennings, M. D. (1998). Large-area mapping of biodiversity. *Annals of the Missouri Botanical Garden*, *85*, 34–47.

Smith, J. H., Stehman, S. V., Wickham, J. D., & Yang, L. (2003). Effects of landscape characteristics on land-cover class accuracy. *Remote Sensing of Environment*, *84*, 342–349.

Smith, J. H., Wickham, J. D., Stehman, S. V., & Yang, L. (2002). Impacts of patch size and land cover heterogeneity on thematic image classification accuracy. *Photogrammetric Engineering and Remote Sensing*, *68*, 65–70.

Stehman, S. V. (1996). Use of auxiliary data to improve the precision of estimators of thematic map accuracy. *Remote Sensing of Environment*, *58*, 169–176.

Stehman, S. V. (2000). Practical implications of design-based sampling inference for thematic map accuracy assessment. *Remote Sensing of Environment*, *72*, 35–45.

Stehman, S. V. (2001). Statistical rigor and practical utility in thematic map accuracy assessment. *Photogrammetric Engineering and Remote Sensing*, *67*, 727–734.

Stehman, S. V., & Czaplewski, R. L. (1998). Design and analysis for thematic map accuracy assessment: Fundamental principles. *Remote Sensing of Environment*, *64*, 331–344.

Stehman, S. V., Wickham, J. D., Smith, J. H., & Yang, L. (2003). Thematic accuracy of the 1992 National Land-Cover Data for the eastern United States: Statistical methodology and regional results. *Remote Sensing of Environment*, *86*, 500–516.

Story, M., & Congalton, R. G. (1986). Accuracy assessment: A user's perspective. *Photogrammetric Engineering and Remote Sensing*, *52*, 397–399.

Verbyla, D. L., & Hammond, T. O. (1995). Conservative bias in classification accuracy assessment due to pixel-by-pixel comparison of classified images with reference grids. *International Journal of Remote Sensing*, *16*, 581–587.

Vogelmann, J. E., Howard, S. M., Yang, L., Larson, C. R., Wylie, B. K., & Van Driel, N. (2001). Completion of the 1990s National Land Cover Data Set for the Conterminous United States from Landsat Thematic Mapper data and ancillary data sources. *Photogrammetric Engineering and Remote Sensing*, *67*, 650–662.

Vogelmann, J. E., Seevers, P. M., & Oimoen, M. (1997). Effects of selected variables for discriminating land cover: Multiseasonal data, different clustering algorithms, and varying numbers of clusters. *Proceedings of the Pecora 13 Symposium, Sioux Falls, SD, August, 1996, CD-ROM.*

Vogelmann, J. E., Sohl, T., Campbell, P. V., & Shaw, D. M. (1998a). Regional land cover characterization using Landsat Thematic Mapper data and ancillary data sources. *Environmental Monitoring and Assessment*, *51*, 415–428.

Vogelmann, J. E., Sohl, T., & Howard, S. M. (1998b). Regional characterization of land cover using multiple sources of data. *Photogrammetric Engineering and Remote Sensing*, *64*, 45–57.

Yang, L., Stehman, S. V., Smith, J. H., & Wickham, J. D. (2001). Thematic accuracy of MRLC land cover for the eastern United States. *Remote Sensing of Environment*, *76*, 418–422.